Monte Carlo Astros

Ethan Triem

2024-11-21

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(Lahman)
library(patchwork)
library(dplyr)
library(ggcorrplot)
```

```
## Warning: package 'ggcorrplot' was built under R version 4.4.2
```

```
library(car)
```

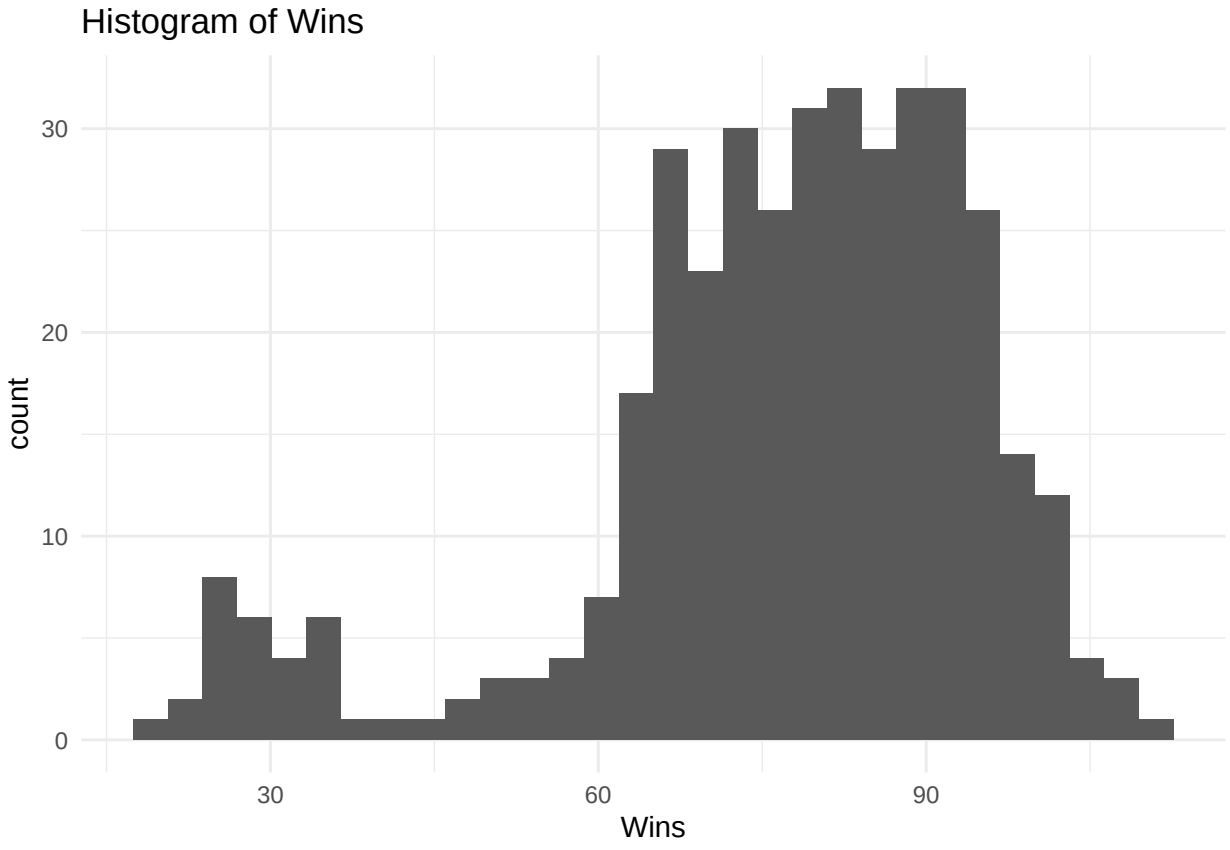
```
## Warning: package 'car' was built under R version 4.4.2

## Loading required package: carData

## Warning: package 'carData' was built under R version 4.4.2

##
## Attaching package: 'carData'
##
## The following object is masked from 'package:Lahman':
##
##     Salaries
##
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
```

```
##
##     recode
##
## The following object is masked from 'package:purrr':
##
##     some
```

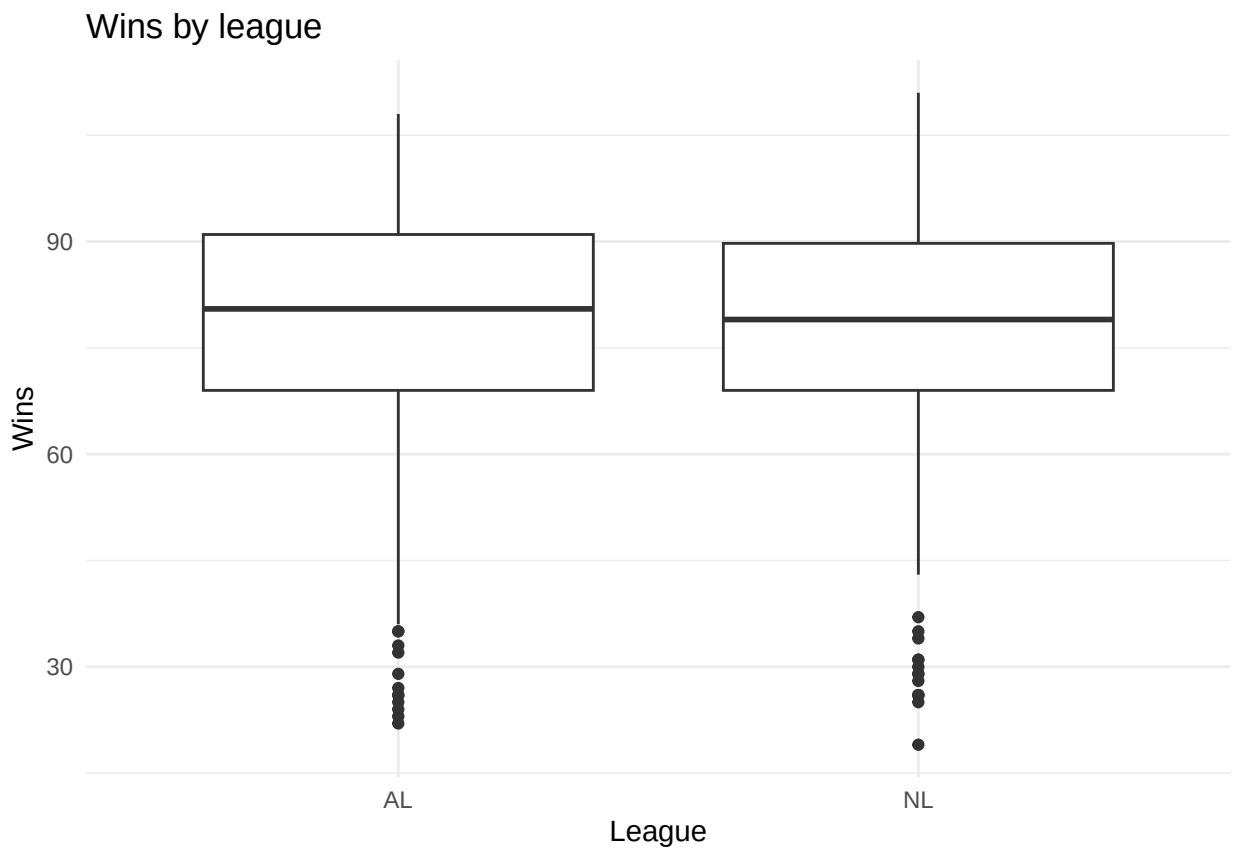
```
teams_2010 <- Teams|> filter(yearID >= 2010)
teams_2010_numeric <- teams_2010[, sapply(teams_2010, is.numeric)]

teams_2010_numeric <- mutate(teams_2010_numeric, X1B = H - X2B-X3B-HR,
                  TB = X1B +2*X2B + 3*X3B + 4*HR,
                  RC = (H + BB)*TB/(AB + BB))

teams_2010_numeric <- mutate(teams_2010_numeric, OBP = (H + BB)/(AB + BB),
                  SLG = (X1B + 2*X2B + 3*X3B + 4*HR)/AB,
                  OPS = OBP + SLG)
```

```
ggplot(data = teams_2010_numeric, aes(x = W)) +
  geom_histogram() +
  theme_minimal() +
  labs(x = "Wins", title = "Histogram of Wins")
```

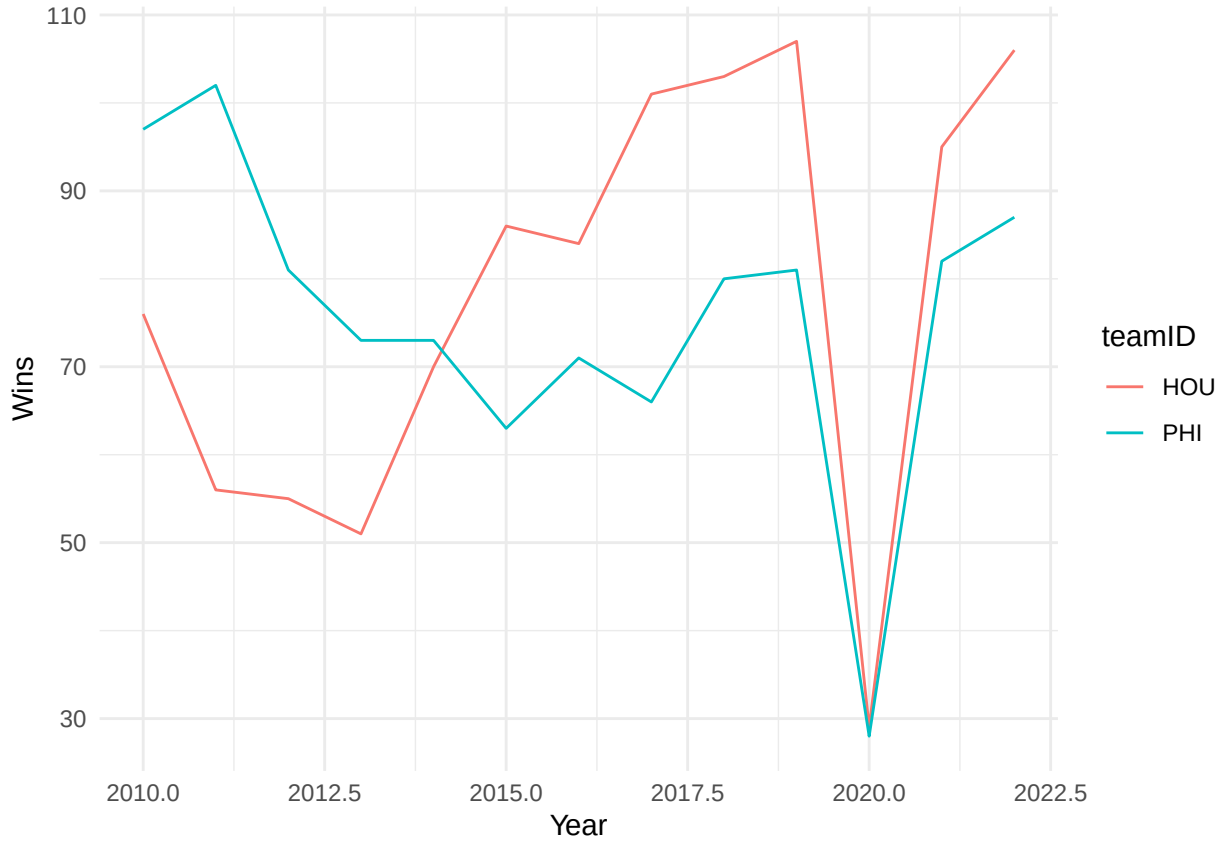
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



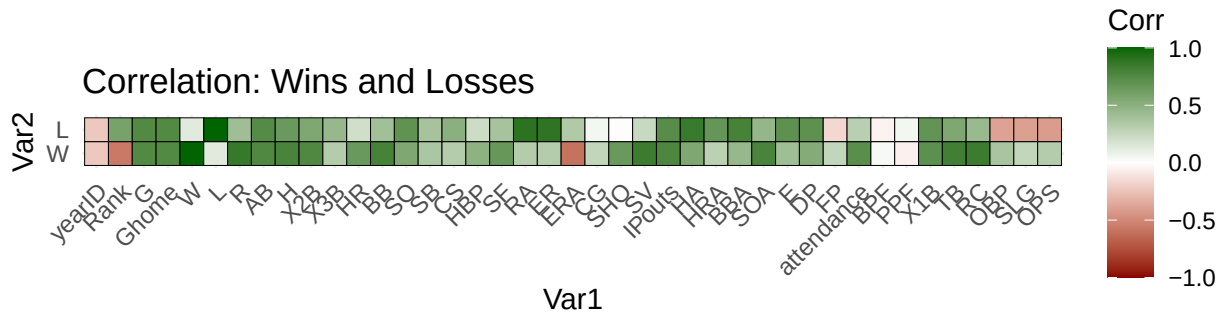
```
ggplot(data = teams_2010, aes(x=lgID, y = W)) +
  geom_boxplot() +
  theme_minimal() +
  labs(x = "League", y = "Wins", title = "Wins by league")
```




```
hou_phi_2010 <- teams_2010 |>
  filter(teamID %in% c("HOU", "PHI"))

ggplot(data = hou_phi_2010, aes(x = yearID, y = W, color = teamID)) +
  geom_line() +
  theme_minimal() +
  labs(x = "Year", y = "Wins")
```




```
cor_matrix <- cor(teams_2010_numeric)
wins_losses_corr <- cor_matrix[c("W", "L"), ]

wins_losses_corr <- t(wins_losses_corr)
ggcorrplot(as.matrix(wins_losses_corr),
           title = "Correlation: Wins and Losses",
           colors = c("darkred", "white", "darkgreen"),
           outline.color = "black",) +
  guides(x= guide_axis(angle=45)) +
  theme(axis.title.x = element_blank(),
        axis.title.y = element_blank()) +
  theme_minimal()
```

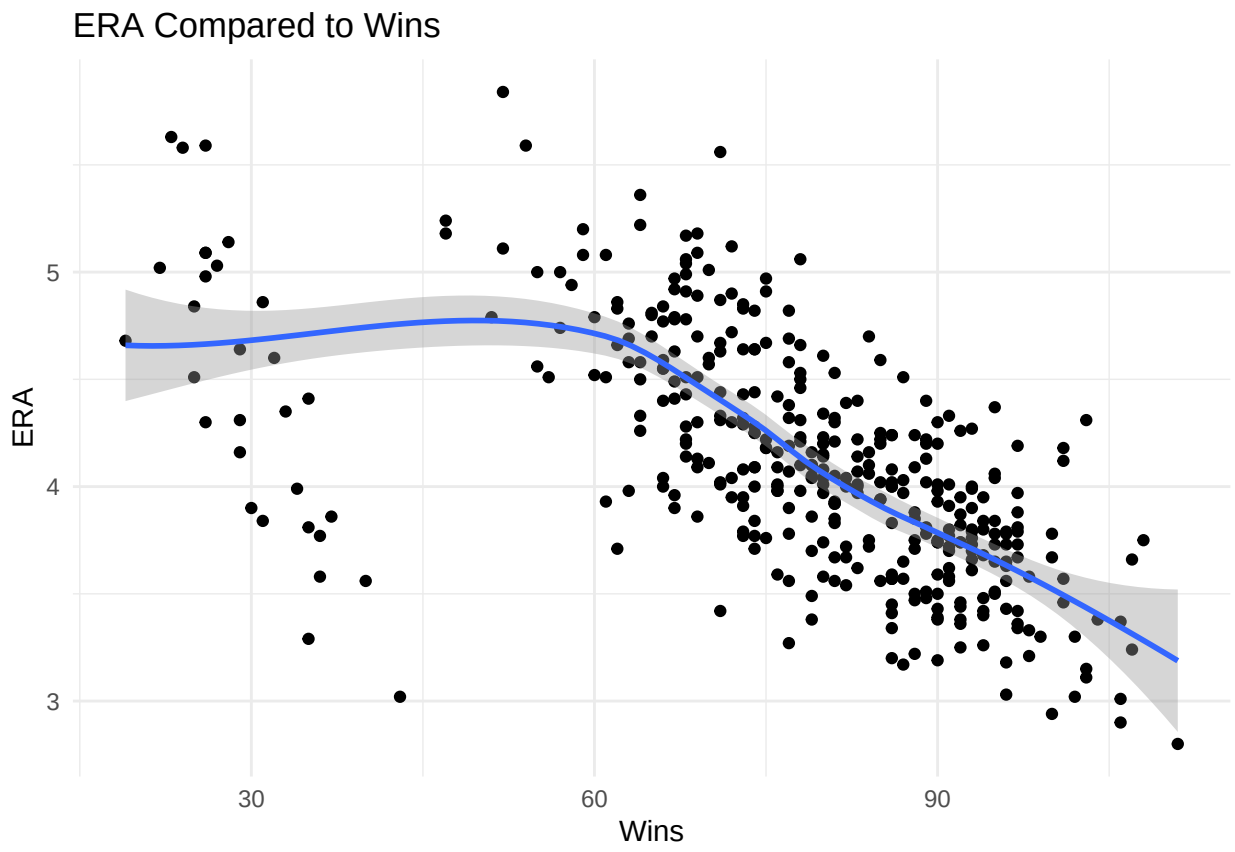



```
wins_corr <- cor_matrix["W", ]
wins_corr_ranked <- sort(wins_corr, decreasing = TRUE)

print(wins_corr_ranked)
```

```
##           W           R          SV          RC          TB          BB
##  1.00000000  0.85594313  0.83640306  0.83575051  0.82108730  0.80231617
##         SOA           H      IPouts          X2B          AB           G
##  0.78838266  0.78530788  0.76877724  0.75980192  0.75784883  0.75225293
##       Ghome  attendance         X1B          SF         SHO          HR
##  0.75082879  0.71786954  0.70636675  0.66114759  0.64931726  0.64786621
##          SO          HA          DP         HBP         BBA           E
##  0.56410086  0.55799724  0.52968511  0.49343192  0.44556675  0.41185070
##         OBP          SB         X3B          RA         OPS          ER
##  0.37397603  0.35688275  0.32236166  0.31996970  0.31795512  0.31784785
##          CS         HRA         SLG          CG          FP           L
##  0.31722963  0.28790198  0.26425455  0.26424562  0.25686102  0.13068065
##         BPF         PPF      yearID        Rank         ERA
##  0.04268864 -0.05747376 -0.23209916 -0.56807003 -0.60337574
```

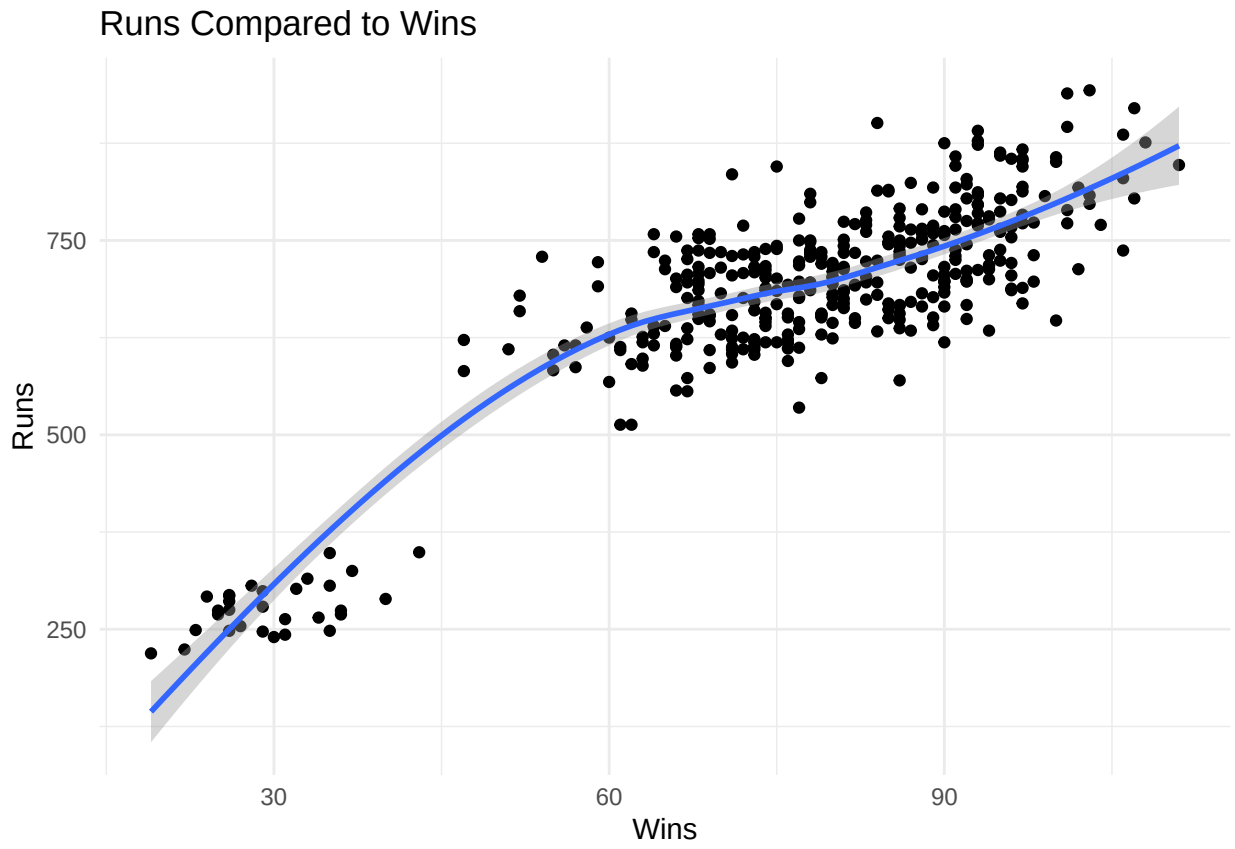
```
ggplot(data = teams_2010_numeric, aes(x=W, y=ERA)) +
  geom_point() +
  geom_smooth() +
  theme_minimal() +
  labs(x = "Wins", y ="ERA", title = "ERA Compared to Wins")
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
ggplot(data = teams_2010_numeric, aes(x=W, y=R)) +
  geom_point() +
  geom_smooth() +
  theme_minimal() +
  labs(x = "Wins", y ="Runs", title = "Runs Compared to Wins")
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
p_vals <- c()

for (var in colnames((teams_2010_numeric))) {
  if (var !="W"){
    formula <- as.formula(paste("W~", var))
    fit <- lm(formula, data = teams_2010_numeric)

    p_val <-summary(fit)$coefficients[2,4]
    p_vals[var] <- p_val
  }
}

p_vals_ranked <- sort(p_vals, decreasing = FALSE)
print(p_vals_ranked)
```

```
##             R            SV            RC            TB            BB
## 3.400397e-113 2.325577e-103 4.702418e-103  1.617615e-96  5.648025e-89
##           SOA             H       IPouts           X2B            AB
##  6.983954e-84  8.244908e-83  2.461494e-77  1.500358e-74  5.835117e-74
##             G         Ghome    attendance           X1B            SF
##  2.663560e-72  6.929530e-72  5.315555e-63  3.414751e-60  2.326865e-50
##           SHO            HR           ERA          Rank            SO
##  4.609785e-48  8.677316e-48  4.865878e-40  1.051045e-34  3.817524e-34
##            HA            DP           HBP           BBA             E
##  2.682664e-33  1.369698e-29  2.525421e-25  2.032851e-20  2.122398e-17
```

```
##           OBP            SB           X3B            RA           OPS
##  2.159757e-14  3.705934e-13  7.003753e-11  9.834617e-11  1.305922e-10
##            ER            CS           HRA           SLG            CG
##  1.325713e-10  1.445589e-10  7.021475e-09  1.181751e-07  1.182952e-07
##            FP        yearID             L           PPF           BPF
##  2.705933e-07  3.616484e-06  9.779211e-03  2.575047e-01  4.005102e-01
```

```
off <- c("R" = .3, "RC" = .25, "TB" = .2, "BB" = .15, "H" = .1)
def <- c("SV" = .3, "SOA" = .25, "IPouts" = .2, "SHO" = .15, "ERA" = .1)

world_series_2022 <- SeriesPost %>%
  filter(yearID == 2022, round == "WS")

teams_2022 <- c(world_series_2022$teamIDwinner, world_series_2022$teamIDloser)

world_series_teams <- Teams %>%
  filter(yearID == 2022, teamID %in% teams_2022)


simulate_game <- function(team1_stats, team2_stats, off, def, n = 10000) {
  team1_off <- sum(team1_stats[names(off)] * off)
  team2_off <- sum(team2_stats[names(off)] * off)

  team1_def <- sum(team1_stats[names(def)] * def)
  team2_def <- sum(team2_stats[names(def)] * def)

  team1_score <- rnorm(n, mean = team1_off / sqrt(162), sd = team1_def / sqrt(162))
  team2_score <- rnorm(n, mean = team2_off / sqrt(162), sd = team2_def / sqrt(162))
  sum(team1_score > team2_score) / n
}

hou_stats <- world_series_teams %>% filter(teamID == "HOU")
phi_stats <- world_series_teams %>% filter(teamID == "PHI")

hou_stats <- mutate(hou_stats,
                    X1B = H - X2B-X3B-HR,
                    TB = X1B +2*X2B + 3*X3B + 4*HR,
                    RC = (H + BB)*TB/(AB + BB))

phi_stats <- mutate(phi_stats,
                    X1B = H - X2B-X3B-HR,
                    TB = X1B +2*X2B + 3*X3B + 4*HR,
                    RC = (H + BB)*TB/(AB + BB))
```

```
set.seed(1)
hou_win_prob <- simulate_game(hou_stats, phi_stats, off, def)
phi_win_prob <- 1 - hou_win_prob

cat("Game 1", "\n")
```

```
## Game 1
```

```r
cat("Astros Win Probability:", hou_win_prob, "\n")
```

```
## Astros Win Probability: 0.4964
```

```r
cat("Phillies Win Probability:", phi_win_prob, "\n")
```

```
## Phillies Win Probability: 0.5036
```

```r
cat("\n")
```

```r
set.seed(2)
hou_win_prob <- simulate_game(hou_stats, phi_stats, off, def)
phi_win_prob <- 1 - hou_win_prob

cat("Game 2", "\n")
```

```
## Game 2
```

```r
cat("Astros Win Probability:", hou_win_prob, "\n")
```

```
## Astros Win Probability: 0.5036
```

```r
cat("Phillies Win Probability:", phi_win_prob, "\n")
```

```
## Phillies Win Probability: 0.4964
```

```r
cat("\n")
```

```r
set.seed(3)
hou_win_prob <- simulate_game(hou_stats, phi_stats, off, def)
phi_win_prob <- 1 - hou_win_prob

cat("Game 3", "\n")
```

```
## Game 3
```

```r
cat("Astros Win Probability:", hou_win_prob, "\n")
```

```
## Astros Win Probability: 0.5029
```

```r
cat("Phillies Win Probability:", phi_win_prob, "\n")
```

```
## Phillies Win Probability: 0.4971
```

```r
cat("\n")
```

```r
set.seed(4)
hou_win_prob <- simulate_game(hou_stats, phi_stats, off, def)
phi_win_prob <- 1 - hou_win_prob

cat("Game 4", "\n")
```

```
## Game 4
```

```r
cat("Astros Win Probability:", hou_win_prob, "\n")
```

```
## Astros Win Probability: 0.4997
```

```r
cat("Phillies Win Probability:", phi_win_prob, "\n")
```

```
## Phillies Win Probability: 0.5003
```

```r
cat("\n")
```

```r
set.seed(5)
hou_win_prob <- simulate_game(hou_stats, phi_stats, off, def)
phi_win_prob <- 1 - hou_win_prob

cat("Game 5", "\n")
```

```
## Game 5
```

```r
cat("Astros Win Probability:", hou_win_prob, "\n")
```

```
## Astros Win Probability: 0.502
```

```r
cat("Phillies Win Probability:", phi_win_prob, "\n")
```

```
## Phillies Win Probability: 0.498
```

```r
cat("\n")
```

```r
set.seed(6)
hou_win_prob <- simulate_game(hou_stats, phi_stats, off, def)
phi_win_prob <- 1 - hou_win_prob

cat("Game 6", "\n")
```

```
## Game 6
```

```r
cat("Astros Win Probability:", hou_win_prob, "\n")
```

```
## Astros Win Probability: 0.5044
```

```r
cat("Phillies Win Probability:", phi_win_prob, "\n")
```

```
## Phillies Win Probability: 0.4956
```

```r
cat("\n")
```

```r
set.seed(7)
hou_win_prob <- simulate_game(hou_stats, phi_stats, off, def)
phi_win_prob <- 1 - hou_win_prob

cat("Game 7", "\n")
```

```
## Game 7
```

```r
cat("Astros Win Probability:", hou_win_prob, "\n")
```

```
## Astros Win Probability: 0.4974
```

```r
cat("Phillies Win Probability:", phi_win_prob, "\n")
```

```
## Phillies Win Probability: 0.5026
```

```r
set.seed(8)
hou_win_prob <- simulate_game(hou_stats, phi_stats, off, def)
phi_win_prob <- 1 - hou_win_prob

cat("Game 1", "\n")
```

```
## Game 1
```

```r
cat("Astros Win Probability:", hou_win_prob, "\n")
```

```
## Astros Win Probability: 0.4964
```

```r
cat("Phillies Win Probability:", phi_win_prob, "\n")
```

```
## Phillies Win Probability: 0.5036
```

```r
cat("\n")
```

```r
set.seed(9)
hou_win_prob <- simulate_game(hou_stats, phi_stats, off, def)
phi_win_prob <- 1 - hou_win_prob

cat("Game 2", "\n")
```

```
## Game 2
```

```r
cat("Astros Win Probability:", hou_win_prob, "\n")
```

```
## Astros Win Probability: 0.5065
```

```r
cat("Phillies Win Probability:", phi_win_prob, "\n")
```

```
## Phillies Win Probability: 0.4935
```

```r
cat("\n")
```

```r
set.seed(10)
hou_win_prob <- simulate_game(hou_stats, phi_stats, off, def)
phi_win_prob <- 1 - hou_win_prob

cat("Game 3", "\n")
```

```
## Game 3
```

```r
cat("Astros Win Probability:", hou_win_prob, "\n")
```

```
## Astros Win Probability: 0.5051
```

```r
cat("Phillies Win Probability:", phi_win_prob, "\n")
```

```
## Phillies Win Probability: 0.4949
```

```r
cat("\n")
```

```r
set.seed(11)
hou_win_prob <- simulate_game(hou_stats, phi_stats, off, def)
phi_win_prob <- 1 - hou_win_prob

cat("Game 4", "\n")
```

```
## Game 4
```

```r
cat("Astros Win Probability:", hou_win_prob, "\n")
```

```
## Astros Win Probability: 0.5052
```

```r
cat("Phillies Win Probability:", phi_win_prob, "\n")
```

```
## Phillies Win Probability: 0.4948
```

```r
cat("\n")
```

```r
set.seed(12)
hou_win_prob <- simulate_game(hou_stats, phi_stats, off, def)
phi_win_prob <- 1 - hou_win_prob

cat("Game 5", "\n")
```

```
## Game 5
```

```r
cat("Astros Win Probability:", hou_win_prob, "\n")
```

```
## Astros Win Probability: 0.5029
```

```r
cat("Phillies Win Probability:", phi_win_prob, "\n")
```

```
## Phillies Win Probability: 0.4971
```

```r
cat("\n")
```

```r
set.seed(13)
hou_win_prob <- simulate_game(hou_stats, phi_stats, off, def)
phi_win_prob <- 1 - hou_win_prob

cat("Game 6", "\n")
```

```
## Game 6
```

```r
cat("Astros Win Probability:", hou_win_prob, "\n")
```

```
## Astros Win Probability: 0.4937
```

```r
cat("Phillies Win Probability:", phi_win_prob, "\n")
```

```
## Phillies Win Probability: 0.5063
```

```r
cat("\n")
```

```r
set.seed(14)
hou_win_prob <- simulate_game(hou_stats, phi_stats, off, def)
phi_win_prob <- 1 - hou_win_prob

cat("Game 7", "\n")
```

```
## Game 7
```

```r
cat("Astros Win Probability:", hou_win_prob, "\n")
```

```
## Astros Win Probability: 0.4935
```

```r
cat("Phillies Win Probability:", phi_win_prob, "\n")
```

```
## Phillies Win Probability: 0.5065
```